APPENDIX G

UNCERTAINTY ANALYSIS ALTERNATIVES

APPENDIX G

Uncertainty Analysis Alternatives Abstract

The objective of this appendix is to describe the uncertainty analysis alternatives for use in the System Assessment Capability (SAC), Rev. 0. The uncertainty approach will provide a framework within which all of the other models in the SAC will operate (Figure G-i).

Output from this activity will provide insight into the variability of the output for the overall results of the assessment and to allow feasibility testing of the SAC (Rev. 0). Information on the overall variability of results will be used to put the results in the context of "how well do we know" the outcome. This information will be described in the risk characterization report to be prepared at the conclusion of the SAC Rev. 0 analysis. The results of feasibility testing will provide insight needed for the design of SAC (Rev. 1). Key aspects of the approach to be tested include computation time, data storage requirements, and methods for generation of stochastic inputs that are compatible with calibrated models.

The proposed uncertainty approach is to link the suite of inventory, transport, and impact models in a stochastic framework. This will require that parameter distribution functions be provided for key model parameters. Special sampling techniques will be employed to reduce computation time. Release and transport calculations will be stored for later use by risk and impact models. Risk and impact calculations will be stored when appropriate for use in calculation of different exposure scenarios.

Figure G-i. System Assessment Capability System Conceptual Model.

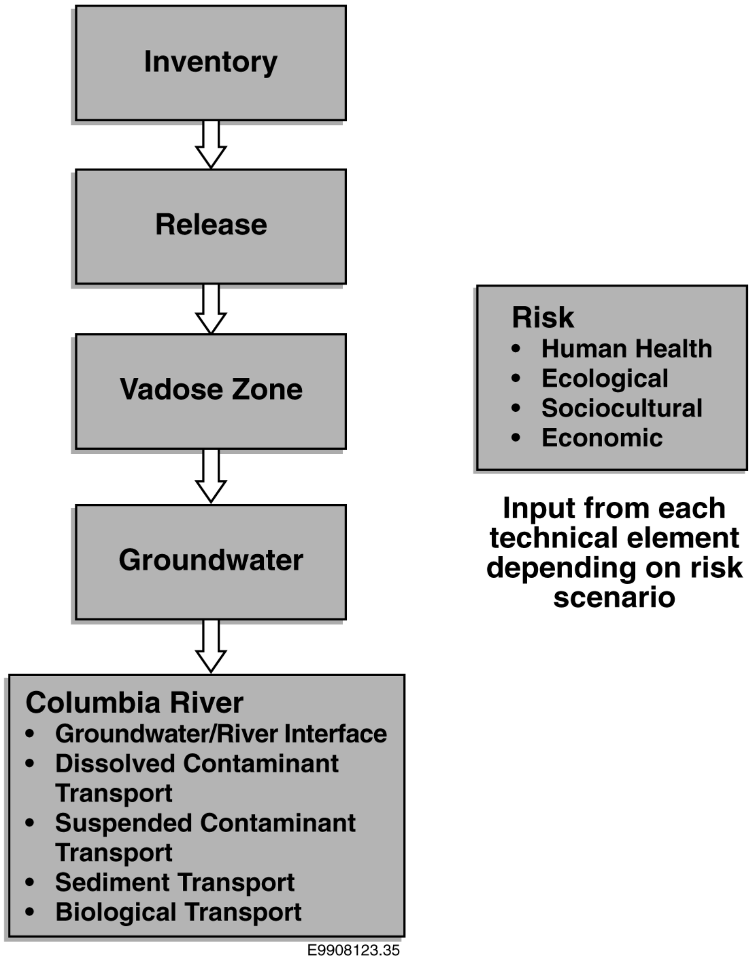


TABLE OF CONTENTS

UNCERTAINTY ANALYSIS ALTERNATIVES

FIGURE

APPENDIX G

UNCERTAINTY ANALYSIS ALTERNATIVES

G.1 BACKGROUND

The System Assessment Capability (SAC) is an evolving capability to assess the cumulative impacts of radioactive and chemical waste at the Hanford Site. The initial assessment (SAC, Rev. 0) is being designed to:

- Examine radioactive and hazardous chemical contaminants that are expected to be dominant contributors to risk and impacts.
- Determine the long-term migration and fate of contaminants from Hanford Site operational areas.
- Include a broad suite of quantitative and qualitative risk and impact metrics.
- Include a quantification of uncertainty.

This appendix addresses the approach to be used in the uncertainty task of the SAC. The overall role of the uncertainty task is to quantify the variability in performance metrics. Possible sources of variability include model parameter variations, conceptual model choices, future states of nature, and different future uses of the Hanford Site. The relationship of uncertainty-related information to the high-level data flow between other conceptual models is provided in Figure 1-1. The uncertainty approach provides a framework within which all of the other models will operate.

G.2 GENERAL CONCEPTS FOR UNCERTAINTY ANALYSES

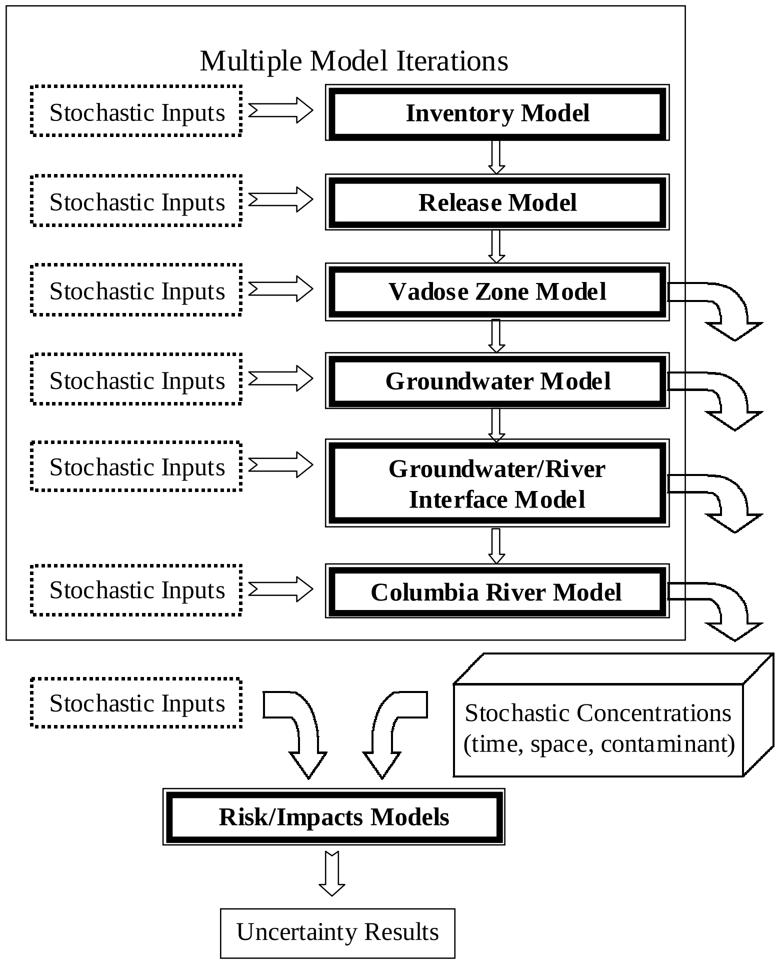
In a general sense, an uncertainty analysis takes a set of stochastic input parameters, passes them through a model or transfer function, and then attempts to obtain the statistical distribution of the resulting outputs. This output distribution can be used to make general inferences, such as the following:

- Describing the range of potential outputs of the system
- Estimating the probability that the output will exceed a specific threshold.

G.2.1 Typical Computation Schemes for Uncertainty Analyses

A number of computation techniques have been used in uncertainty analyses for large-scale modeling projects. A brief overview of the most common techniques is provided here.

Figure G-1. Relationship of Uncertainty-Related Information to other Conceptual Models for the SAC (Rev. 0).



- **Analytical Propagation**: When there are few stochastic input parameters, and when the model is not too complicated, one can sometimes obtain an analytic form for the output statistical distribution. Several variations of this general approach have been advanced, but the problem of obtaining an analytical solution is intractable if the model describes a process such as the transport of contaminants through an unsaturated zone.

- **Monte Carlo with Random Sampling**: There are two basic steps in a Monte Carlo approach using random sampling. One generates a value for every stochastic parameter in the model and then executes the model, thereby obtaining an output value. This process is often called one realization. One then repeats the entire process, obtaining another output that is different from the first, but which is as equally likely to occur as the first output. After repeating this process a number of times, one has a set of equally likely outputs that represent the statistical distribution of all outputs. Although conceptually simple, a shortcoming is that many realizations may be required to obtain a satisfactory description of the output distribution.

- **Monte Carlo with Specialized Sampling Techniques**: Several specialized sampling techniques have been developed to reduce the number of realizations required in a Monte Carlo analysis to obtain a satisfactory description of the output distribution. One of the techniques, called Latin Hypercube Sampling (Iman and Conover 1982), has proven to be successful for mass transport applications in groundwater systems. The general approach is the same as for Monte Carlo modeling with random sampling, but the specific values of the input parameters are chosen differently. They are chosen from the same statistical distributions, but the sampling scheme spreads the values in such a way as to reduce sampling variability.

G.2.2 What an Uncertainty Analysis Can Provide

The result of a numerically-based uncertainty analysis is a set of output values and the probability of occurrence for each value. These values are representative of the statistical distribution of the system outputs. The most common uses for uncertainty analyses are to make general inferences, such as the following:

- Estimating the mean value of the outputs, and stating the standard deviation of the mean
- Estimating the probability the output will exceed a specific threshold
- Putting a confidence interval on a function of the outputs
- Describing the complete range of potential outputs of the system.

Implicit in this approach is the assumption that statistical distributions for the input parameters were correctly chosen to model reality, and the assumption that the model is a realistic description of the dominant features, events, and processes of the system. Since neither of these assumptions are likely to be entirely correct, a complete interpretation of the results should consider deviations from these assumptions.

The first assumption is easy to check: one simply uses different statistical distributions for the input parameters. If the outputs vary significantly compared to the use of the original distributions, then the analyst must take additional care to support specification of the input distributions. The assumption of a correct model choice can be supported two ways. The first way is to obtain consensus among interested parties that the correct model has been chosen. The second way is to implement alternative conceptual models, and then run the Monte Carlo analysis again. There is a formal statistical approach to handle multiple conceptual models, but it is rarely used because it requires one to develop an entire suite of plausible conceptual models.

An uncertainty analysis is sometimes confused with a sensitivity analysis. In a general sense, an uncertainty analysis attempts to describe the entire set of possible outcomes of the model, together with their associated probability of occurrence. A sensitivity analysis, however, attempts to determine the amount the model output will change given small changes in model inputs. A sensitivity analysis thus measures the change in the model in a localized region of the space of inputs. However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses.

G.3 HISTORICAL USE OF UNCERTAINTY ANALYSES FOR LARGE PROJECTS

This section contains a brief overview of the ways uncertainty analyses have been conducted for major projects. The order of appearance for project descriptions has no particular significance.

- **Waste Isolation Pilot Plant**: The Waste Isolation Pilot Plant (WIPP) recently received a license for the U.S. Environmental Protection Agency (EPA) to begin transuranic waste acceptance operations as a repository. The license application utilized a linked suite of complex numerical models and embedded those models in a Monte Carlo framework that used 100 realizations. The project utilized both Latin Hypercube sampling and special combinations of intermediate results to build a suite of complementary cumulative distribution functions for potential dose to humans. The primary compliance criterion for the uncertainty analysis was an exceedance probability (DOE 1996a).

- **Yucca Mountain Project**: The U.S. Department of Energy (DOE) is studying Yucca Mountain, Nevada, as the potential site for a high-level radioactive waste repository. The recently published Viability Assessment (DOE 1998a) contained an uncertainty analysis for dose to humans in the future at a variety of distances from the repository. The modeling approach was a combination of linked complex models and response surface data generated by other complex models. The project utilized Latin Hypercube sampling in a Monte Carlo framework and presented results that were the mean value and the 95th percentile of potential dose to humans, based on 100 model realizations.

- **Hanford Environmental Dose Reconstruction Project**: The Hanford Environmental Dose Reconstruction (HEDR) project estimated historical doses to individuals living around the Hanford site. The modeling approach was a linked suite of complex models at a limited set of output locations. The linked suite of models produced a data set of environmental concentrations that could be quickly accessed to compute dose to an individual. Results included maps of median exposures for a variety of types of individuals, and suites of 100 dose estimates for specific individuals (Farris et al. 1994).

- ***Columbia River Comprehensive Impact Assessment* (Human Risk)**: Human risk for the *Columbia River Comprehensive Impact Assessment* (CRCIA) used a fully stochastic model. This model used over 50 contaminants at 27 locations and accessed a suite of stochastic concentration data generated from historical sampling data. Calculated results included 100 realizations of dose for a variety of river use scenarios (DOE 1998b).

- ***Columbia River Comprehensive Impact Assessment* (Ecological Risk)**: Ecological risk for the CRCIA used a fully stochastic food-web based ecological model. This model used over 50 contaminants and 50 species at 27 locations and accessed a suite of stochastic concentration data generated from historical sampling data. Calculated results included 100 realizations of body burdens and dose to individual organisms (DOE 1998b).

- **Retrieval Performance Evaluation Project**: The Retrieval Performance Evaluation Project studied past and possible future leaks and residual tank waste in the AX Tank Farm to develop methodologies and identify data needs required to support future cleanup decisions. A stochastic analysis was conducted on a subset of the project models, using linked simplified analytical models. Stochastic results reported included cumulative probabilities, range, and mean value of risk to humans from the tank inventory for a variety of scenarios (DOE 1999a).

- **Hanford Remedial Action Environmental Impact Statement and Comprehensive Land Use Plan**: A unit risk factor approach was used to analyze over 200 contaminants for the draft *Hanford Remedial Action Environmental Impact Statement*. This simplified approach provided stochastic results, but some steady-state assumptions were utilized to generate the unit transport and risk factors (DOE 1996b). This approach was dropped in the revised draft of this document, due in part to a change of scope (DOE 1999b).

- **Analytica Application**: The tracking and analysis framework model provides support to the National Acid Precipitation Assessment Program. The tracking and analysis framework model is implemented in Analytica®, which is a software tool produced by Lumina Decision Systems. Analytica is a visual software tool for creating, analyzing, and communicating quantitative models. The tracking and analysis framework model components are based upon reduced form models (or transfer matrices) derived from more detailed scientific models. Distributions are propagated using conventional Monte Carlo or Latin Hypercube sampling methods (Henrion et al. 1997).

- **Hazardous Waste Identification Rule**. The purpose of the Hazardous Waste Identification Rule assessment is to develop standards for chemical concentrations in hazardous waste before disposal. Meeting the standards would allow the waste to exit the hazardous waste category under the *Resource Conservation and Recovery Act of 1976* (RCRA), Subtitle C, and be treated as industrial waste under the *Act*, Subtitle D. 17 empirical and semi-analytical models were linked together in the FRAMES software, using a Monte Carlo approach. The resulting outputs can be used to assess the probability of a concentration meeting specified levels of human risk (Whelan and Laniak 1998).

G.3.1 Major Steps in Parametric Uncertainty Analyses

Uncertainty in the risk conceptual model predictions can arise from a number of sources, including specification of the problem, formulation of conceptual model, formulation of the computational model, estimation of parameter values, and calculation, interpretation, and documentation of results (BIOMOVS 1993). Of these sources, only uncertainties due to estimation of parameter values can be quantified in a straightforward manner. The main steps in parameter uncertainty analysis are as follows:

1. Identify the parameter that could contribute significantly to the uncertainty in the final model prediction.

2. Construct a probability density function (PDF) for each parameter to reflect the belief that the parameter will take on various values within its possible range.

3. Account for dependency (correlation) among the parameters.

4. Propagate the uncertainties through the model to generate a PDF of predicted values.

5. Derive confidence limits and intervals from the PDF of predicted values to provide a quantitative statement about the effect of parameter uncertainty on the model prediction.

G.4 POSSIBLE UNCERTAINTY ANALYSIS APPROACHES FOR SYSTEM ASSESSMENT CAPABILITY (REV. 0)

The proposed uncertainty approach is a linked suite of fate and transport models in a stochastic framework. The stochastic framework would utilize specialized sampling techniques (such as Latin Hypercube sampling) to optimize computation time. Once this high-level approach is chosen, additional details are tightly integrated with the overall computational framework.

Sometimes qualitative statements concerning uncertainty can be made for many of the submodels in a system model. However, it is usually impossible to translate these qualitative statements about a submodel into qualitative or quantitative statements about the output of the entire system. In addition, the system must provide quantitative risk or impact results. A quantitative uncertainty statement would be desired to support the quantitative risk or impact statements. For these reasons, the proposed conceptual model for the uncertainty analysis is quantitative in approach.

For the SAC, Rev. 0, a variety of performance measures must be supported by uncertainty analyses. These performance measures are in the broad areas of human risk, ecological risk, cultural impacts, and socio-economic impacts. These performance measures must be supported for a wide variety of locations on the Hanford Site, and along the Columbia River.

G.4.1 Computational Constraints

Because quantitative uncertainty estimates require multiple model runs, there are a number of areas where design constraints must be carefully addressed. Three major design considerations are discussed briefly in this section: computation time, data storage, and utilization of stochastic inputs.

- **Computation Time**: Because the system model may be run 100 times for several Hanford Site future use scenarios, computation time of the submodels is a major concern. A computation time budget must be established for the overall system, and for each submodel. This computation time budget should be based on a project decision of how long a user is willing to wait to obtain a particular type of answer from the system. Design tradeoffs to comply with a time budget include choices on complexity of submodels, computer systems, and the use of parallel computing techniques.

- **Data Storage**: Results from the uncertainty analysis are desired for a variety of risk metrics, many geographic locations, many times, and a substantial number of contaminants. The large number of results can result in very large data storage requirements. Design tradeoffs to comply with a disk storage budget include issues such as limiting the number of locations and times steps at which impacts will be calculated and designing model repeatability. If results are repeatable, they may be recalculated if desired (rather than being computed once and stored).

- **Utilizing Stochastic Values**: Fate and transport models may require calibration in order to achieve some degree of history matching on contaminant migration. It is a substantial technical challenge to generate stochastic inputs that are compatible with calibrated models. Proper conditioning of stochastic analyses with observed data may be the most difficult conceptual model issue to be addressed in system design studies.

G.4.2 Qualitative Uncertainty Considerations

If it is not possible to apply formal Monte Carlo analysis in a reasonable manner because not enough is known about the underlying relationships, it may still be possible to provide a quantitative or semi-quantitative rating of the risk issues that may be useful in some contexts. The following summary, adapted from guidance currently being used by the Intergovernmental Panel on Climate Change, shows one possible such method for rating the confidence in the underlying science.

1. For each of the major findings you expect to develop, **identify the most important factors and uncertainties that are likely to affect the conclusions**. Also specify which important factors/variables are being treated exogenously or fixed, as it will almost always be the case that some important components will be treated in this way when addressing complex phenomena.

2. **Document ranges and distributions in the literature**, including sources of information on the key causes of uncertainty. Note that it is important to consider the types of evidence available to support a finding, such as distinguishing findings that are well established through observations and tested theory from those that are not so well established.

3. Given the nature of the uncertainties and the state of science, **make an initial determination of the appropriate level of precision**—is the state of science such that only qualitative estimates are possible, or is quantification possible, and if so, to how many significant digits?

As the assessment proceeds, recalibrate the level of precision in response to your assessment of new information.

4. Quantitatively or qualitatively **characterize the distribution of values that a parameter, variable, or outcome may take**. First identify the end points of the range, and/or any high consequence, low probability outcomes or "outliers." Particular care needs to be taken to specify what portion of the range is included in the estimate (for example, this is a 90% confidence interval) and what the range is based on. Then provide an assessment of the general shape (for example, uniform, bell, bimodal, skewed, symmetric) of the distribution. Finally, provide your assessment of the central tendency of the distribution (if appropriate).

5. **Rate and describe the state of scientific information** upon which the conclusions and/or estimates (that is from Step 4) are based.

6. **Prepare a "traceable account"** of how the estimates were constructed that describes the reasons for adopting a particular probability distribution, including important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, and critical uncertainties.

7. OPTIONAL: **Use formal probabilistic frameworks for assessing expert judgment** (that is decision-analytic techniques), as appropriate.

In describing the state of scientific information, the following descriptors may be appropriate:

- **Well-Established**: models incorporate known processes; observations consistent with models; or multiple lines of evidence support the finding).

- **Established but Incomplete**: models incorporate most known processes, although some parameterizations may not be well tested; observations are somewhat consistent but incomplete; current empirical estimates are well founded, but the possibility of changes in governing processes over time is considerable; or only one or a few lines of evidence support the finding.

- **Competing Explanations**: different model representations account for different aspects of observations or evidence, or incorporate different aspects of key processes, leading to competing explanations.

- **Speculative**: conceptually plausible ideas that have not received much attention in the literature or that are laced with difficult to reduce uncertainties.

G.4.3 Alternative Computational Approaches

Three major computational approaches for uncertainty analyses are provided as possible alternatives:

- **Linked Complex Models**: Figure 1-1 provides a schematic showing component models linked together in a stochastic framework. The linked models could be complex models of each system component. The computation time and data storage issues addressed in Section 4.1 become major design considerations in this approach. However, this approach has been taken in a regulatory setting for the Waste Isolation Pilot Plant, and also for the proposed high-level repository at Yucca Mountain. This approach was also utilized in the HEDR project. To address the assumption that many different exposure scenarios could be run for a given set of transport calculations, the calculations would be divided into two stages. The first stage would be to calculate and store the estimates of environmental concentrations at multiple locations, multiple times, and for all contaminants. The stored data would be accessed and utilized for risk and impact calculations. This would allow relatively quick computation of risk results after the concentration data set was generated.

- **Linked Simple Models**: The linked models in Figure 1-1 could also be simplified models of each system component. This approach may somewhat alleviate the computation time issue addressed in Section 4.1, but does not address data storage issues. This approach has often been utilized for environmental impact statements and screening analyses. This approach could also use the two-stage computational scheme described for linked complex models.

- **Reduced-Form Models**: Another approach is to replace the component models in Figure 1-1 with reduced-form models. Then, one could possibly compute a new value every time a result is desired, rather than storing large amounts of intermediate data. The benefits from this approach are computational speed and greatly reduced data storage requirements. A drawback of this approach is that the complex models must be run many times to develop the reduced form models (often implemented as a convolution or response surface). One then often loses the ability to examine the modeling sequence for the physical meaning behind a particular result of interest. Another drawback of this approach is that convolutions and response surfaces are smoothed functions, and may underestimate the variability computed by more complex models.

G.5 SUMMARY AND RECOMMENDATIONS

The proposed SAC (Rev. 0) approach for uncertainty involves a Monte Carlo approach that has the following major attributes:

- Specialized sampling techniques that would be employed to reduce computation time.

- Complex or moderately complex models that would be linked together into a system model.

- Release and transport calculations that would be conducted, with the results stored for later use by risk and impact models.

- Risk and impact calculations that would utilize stored results to allow quick calculation of different exposure scenarios.

Although a proposed approach for uncertainty calculations is provided, no recommendation for the specific architecture, stochastic modules, and simulation control software is made at this time. Choice of these items would be deferred until initiation of a software design phase.

G.6 REFERENCES

BIOMOVS, 1993, *Biospheric Model Validation Study, Phase II, Technical Report No. 1 Guidelines for Uncertainty Analysis*, Swedish Radiation Protection Institute, Stockholm, Sweden.

DOE, 1996a, *Title 40 CFR 191 Compliance Certification Application for the Waste Isolation Pilot Plant*, DOE/CAO-2184, U.S. Department of Energy, Carlsbad, New Mexico.

DOE, 1996b, *Hanford Remedial Action Environmental Impact Statement and Comprehensive Land Use Plan*, DOE/EIS-0222D, Draft, U.S. Department of Energy, Richland, Washington.

DOE, 1998a, *Viability Assessment of a Repository at Yucca Mountain*, DOE/RW-0508, U.S. Department of Energy, Las Vegas, Nevada.

DOE, 1998b, *Screening Assessment and Requirements for a Comprehensive Assessment*, DOE/RL-96-16, Rev. 1, U.S. Department of Energy, Richland, Washington.

DOE, 1999a, *Retrieval Performance Evaluation Methodology for the AX Tank Farm*, DOE/RL-98-72, U.S. Department of Energy, Richland, Washington.

DOE, 1999b, *Hanford Remedial Action Environmental Impact Statement and Comprehensive Land Use Plan*, DOE/EIS-0222D, Revised Draft, U.S. Department of Energy, Richland, Washington.

Farris, W. T., B. A. Napier, P. W. Eslinger, T. A. Ikenberry, D. B. Shipler, and J. C. Simpson, 1994, *Atmospheric Pathway Dosimetry Report, 1944-1992*, PNWD-2228 HEDR, Battelle Pacific Northwest Laboratory, Richland, Washington.

Henrion, M., R. Sonnenblick, and K. S. Hoo, 1997, *Lessons in the Collaborative Construction of an Environmental Integrated Assessment Model: The Tracking and Analysis Framework*, Lumina Decision Systems, Los Gatos, California.

Iman, R. L. and W. J. Conover, 1982, "A Distribution-free Approach to Inducing Rank Correlations Among Input Variables," in *Communications in Statistics* Vol. B11, No. 3, pp. 311-334.

Resource Conservation and Recovery Act of 1976, 42 U.S.C. 6901, et seq., as amended.

Whelan, G. and G. F. Laniak, 1998, "A Risk-Based Approach for a National Assessment," in *Risk-Based Corrective Action and Brownfields Restorations*, C. H. Benson, J. N. Meegoda, R. B. Gilbert, and S. P. Clemence, eds., pp. 55-74, Geotechnical Special Publication Number 82, American Society of Civil Engineers, Reston, Virginia.